Journal of Peptide Science, Vol. 2, 233–239 (1996)

The Use of Proton Chemical Shifts to Define the Solution Structure of a Dimeric Peptide

BERNARD BUSETTA and PHILIPPE PICARD

Laboratoire de Cristallographie, Université Bordeaux I, France

Received 18 December 1995
Accepted 5 February 1996

Abstract: For flexible peptides, nuclear Overhauser Effects (NOE) experiments do not provide enough information to ensure a correct definition of their solution structure. The use of distance constraints, derived from the knowledge of proton chemical shifts, is developed to restrict the number of possible conformations. In the case of flexible molecules, randomization appears as an important factor of the correct estimation of the chemical shifts from the 3D structure. The refinement of the solution structure of the highly flexible AVP-like parallel dimer is described to illustrate this process.

Keywords: NMR; chemical shift estimation; restrained refinementPB-931 067

INTRODUCTION

Conformational parameters (NOE constraints, ^{3}J-coupling constants, temperature coefficients) are extracted from NMR spectroscopy and currently used in restrained molecular energy refinements [1]. Methods based on the metric matrix [2] or a variable target function [3] have been widely applied in the structure determination of proteins. The method have not been so widely applied to the determination of the solution structure of flexible peptides mainly because of the smaller number of constraints in peptides; there are many fewer NOEs per residue when compared to proteins because of the smaller core to surface ratio and a greater flexibility.

Several sources of conformational informations (such as ^{3}J constraints [4] or hydrogen bond information [5]) were proposed to overcome this problem. However, the proper handling of hydrogen bond restraints, is not straight forward since only the donor is independently determined from experiments. Furthermore, hydrogen bonds are very rare in flexible peptides.

Address for correspondence: Dr Bernard Busetta, Laboratoire de Cristallographie, 351, cours de la Libération, 33405 Talence Cedex, France.

CCC 1075-2617/96/040233-07

As carbon (C13) chemical shifts mainly depend on the neighbour atoms of the polypeptidic main chain [6], minimization of the functions which determined the C13 chemical shifts was used to improved protein conformations. The refinement of protein structure against protein chemical shifts was also described [7].

In this paper, we propose another way to use proton chemical shifts to derive a refinement process of flexible peptide structures (in all the following chemical shifts are expressed in ppm).

MATERIALS AND METHODS

Estimation of Proton Chemical Shifts from 3D Structures

Advances in NMR instrumentation and methodology have now made it possible to determine site-specific proton chemical shift assignments for a large number of proteins and then for small peptides. For this purpose, it is useful to decompose contributions to chemical shifts (δ) into local and non-local contributions [8]. Local 'diamagnetic' and 'paramagnetic' contributions (δ_{local}) are approximated by the observed shifts on short peptides that appear to be in random coil conformations. Each proton chemical shift is thus defined for an average chemical environment. More distant parts of the molecule are

responsible for the $\Delta\delta$ chemical shifts from this initial environment; $\Delta\delta$ gives an information on the chemical environment of a given proton compared to its average chemical environment in a peptide with a random 3D structure. It is expressed as a sum of independent terms:

$$\Delta\delta = \delta_{tot} - \delta_{local} = \delta_{cc} + \delta_{m} + \delta_{el} + \delta_{solv} \qquad (1)$$

where δ_{cc} represents the contribution from ring current effects, δ_{m} is the contribution from magnetic anisotropies of peptide and amide groups and δ_{el}, and δ_{solv} are the electrostatic and solvent effects.

The decomposition allows an empirical estimation of the variations of the H_α proton chemical shifts with an accuracy of 0.230–0.250 p.p.m [7]. Solvent effects which are not yet well understood may explain the great discrepancy which is observed in the estimation of the H_N chemical shifts.

Theoretical proton chemical shifts were computed with Eq. (1) using the empirical parameters which were derived from NMR protein structures by Osapay and Case [9]. The main interest of the AVP-like factors comes from the absence within the molecule of any ring currents effects and hydrogen bonding effects which involve generally strong perturbations for the proton chemical shifts of large proteins. In our case the basic Eq. (1) is reduced to magnetic and electrostatic effects.

For small proteins, to ensure a good estimation of the deplacement of the H_N protons which are implicated in an hydrogen bond, the use of an additive term in Eq. (1) is necessary [10]. As suggest previously [11], this extra variation of the chemical shifts may be described by a function in $1/r^3$ (where r is the distance between the oxygen atom and the H_N proton which are linked by the hydrogen bond).

$$\delta = -0.92 + 7.6/r^3$$

This additive term is null for $r = 2.03$ Å which is roughly the average NH...O distance of a HN proton bound to a water molecule in a random peptide.

For non-H-bonded H_N another additive term is necessary to take into account the temperature dependence of the H_N shifts between the experimental condition (θ °C) and the room temperature (36 °C) at which random coil chemical shifts were determined [12].

$$\delta = -0.01 \times (\theta - 36°)$$

Finally the pH effects were corrected for the side-chain protons of the ionic amino acids by refering the observed shift variations of a given experimental pH to the random coil chemical shifts recorded at the same pH value [12]. The extra weak correction of pH effects on other chemical shifts [13] was not used.

In small peptides, a restricted number of contributions are responsible for the variation of the proton chemical shifts; then it is possible to analyse each of them separately. Furthermore, as described in this paper, observed chemical shifts may be used to obtain a more accurate definition of a peptide solution structure. Compared to well-structured proteins, for flexible peptides (as well as for an unstructured loop in proteins) it is necessary to take into account the averaging of the different contributions. Flexibility appears as the main factor altering the basic Eq. 1.

NMR Spectroscopy of AVP-like Dimer

The parallel dimer of an insect arginine-vasopressine AVP-like factor, used in this study was synthesized as described elsewhere (Picard *et al.* to be published). It consists of two nonapeptidic chains with the primary sequence:

Cys-Leu-Ile-Thr-Asn-Cys-Pro-Arg-Gly-NH_2 (1–9)
| |
Cys-Leu-Ile-Thr-Asn-Cys-Pro-Arg-Gly-NH_2 (1′–9′)

which are linked together via two disulphide bridges (1–1′ and 6–6′). The dimer forms a 12-residues membered N terminal ring, with two tripeptide tails.

2D-NMR spectroscopy (COSY, HoHaHa and ROESY spectra) was performed in water on a Bruker AMX-600 spectrometer at 30 °C and pH = 5.2. No reference standard was used, but as the peptide seems highly flexible, a random coil NMR chemical shift was assumed for the methyl protons of both Ile and Leu residues, providing a way to correct the observed chemical shifts, first referenced to the water resonance (4.75 ppm). Both chains are identical from an NMR point of view: a single set of proton chemical shifts is recorded with the result that no information is available to describe the inter-residual contacts between the chains. The proton chemical shifts observed for the AVP-like parallel dimers are reported in Table 1. The H_N protons are shifted down field and may correspond to a quasi extended conformation [14] of the chain as also suggested by the available coupling constants which are between 7 and 9 Hz. The C-terminal Glycine is completely in a random coil conformation as suggested by the observed chemical

Table 1 Observed Chemical Shifts for the Parallel AVP-like Dimer Recorded in H_2O at 30 °C, pH5.2

	HN	Hα	Hβ	Hγ	Others			
CYS1		4.28	3.13, 3.33					
LEU2	8.41	4.52	1.63, 1.63	1.61	Qδ	(0.90, 0.94)		
ILE3	8.41	4.30	1.89	1.21	Qγ2	0.93	Qδ1	0.88
THR4	8.20	4.42	4.23	1.20	Hδ	6.86		
ASN5	8.43	4.78	2.76, 2.84		Hδ22	7.56		
CYS6	8.38	4.85	2.84, 3.08					
PRO7		4.43	1.96, 2.31	2.02, 2.04	Hδ	3.72, 3.78		
ARG8	8.45	4.34	1.81, 1.92	1.68	Hδ	3.24, 3.24		
GLY9	8.38	3.94						

shifts (0.03 p.p.m. as averaged variations of chemical shifts from the random coil state).

From the ROESY spectrum, beside the intra residue constraints, only distance constraints between adjacent residues are recorded (i.e. 32 intra-residue constraints and 30 constraints between adjacent residues). The DIANA program [3] was used to generate 1000 randomly independent conformers, using the observed ROESY constraints. Then the 50 top conformers were refined by 200 cycles of conjugate gradient process (Powell) with the Xplor package [15].

RESULTS AND DISCUSSION

Randomization Effects

In Table 2 are indicated the variation of the chemical shifts from the random values [12]. (For each side-chain only the average of the absolute variation over all its protons is considered.)

The comparison between the absolute variations of the side chain and H_α protons indicates that the vibration of the side chain follows the vibration of the mainchains and vice versa. The formation of disulphide bridges makes the cysteine side-chains more rigid and the peptide mainchain is less flexible at the level of the cysteine residues. Side chains are always more agitated than their main chain. The randomization factor should be greater for side chain protons than for H_α protons.

Randomization of the variation of H_α proton chemical shifts was already observed in the unstructured loops of proteins [14] and was explained by a local equilibrium between helical and extended conformations. In highly flexible peptides the chemical environment of each proton should be described as an equilibrium between a statistical environment which remembers the averaged environment in the random coil state and the specific environment in the peptide 3D structure.

Table 2 Relative Variations of the Different Proton Chemical Shifts from the Observed Chemical Shifts in Random Coil Peptides. For Each Side Chain, the Reported Variation (in p.p.m.) is Averaged over all its Protons.

	⟨|δ| side chain⟩	⟨|$\delta_{H\alpha}$|⟩	δ_{HN}
Cys1	0.105	0.395	
Leu2	0.043	0.139	0.000
Ile2	0.015	0.077	+0.210
Thr4	0.018	0.050	−0.044
Asn5	0.023	0.031	+0.055
Cys6	0.160	0.169	+0.072
Pro7	0.048	0.035	+0.181
Arg8	0.015	0.065	0.000
Gly9		0.030	

The observed variations of the proton chemical shifts partly reflect the structured conformation and may be estimated by an altered Eq. (1).

$$\Delta\delta = (\delta_{rc} + \delta_m + \delta_{el} + \delta_{solv}) \times p \qquad (2)$$

where $p\ (0 < p < 1)$ represents the proportion of conformers which present such a specific environment. In all the following, p will be the randomization factor which gives the best fit between observed and computed variations of the chemical shifts.

In Table 3 are reported for each residue the randomization factor which should be used in Eq. (2) to give the best fit between the observed and calculated chemical shifts. The resulting p values clearly indicate that the dimer is hardly flexible at the level of the disulphide bridges. Then the chemical shifts of cysteine protons should be more relevant in the determination of the solution 3D-structure.

Table 3 Optimum Randomization Factor p for Hα and Side-chain Protons Defined as the p Values Which Should be Used in Eq. (2) to Provide the Best Fit Between the Observed and the Calculated Chemical Shifts.

	H_α	Side-chain
Cys1	0.65	0.60
Leu2		
Ile3	0.35	0.28
Thr4		
Asn5	0.30	0.25
Cys6	0.50	0.45
Pro7	0.50	0.30
Arg8	0.35	0.25

Randomization appears as an important factor for the correct estimation of the chemical shifts of flexible peptides. We look if this observation, performed on a highly flexible peptide, is still valid on proteins. Equation (2) was applied to small protein 3D structures described in the Protein Data Bank (PDB) [16] which were determined from 2D NMR spectroscopy. As for flexible peptides p is the optimum randomization factor for the whole protein. To take into account the greater flexibility of the unstructured loops a weaker randomization factor was used (0.70p) when no hydrogen bond involves their peptide groups. For the different side chains (excepted for SS bridged cysteines) another 0.7 decrease of the randomization factor is applied, compared to the randomization which is used for the main chain).

In Table 4 are reported the overall optimum randomization factor for peptides and proteins of increasing size. It clearly appears that Osapay and Case's formulation to estimate the proton chemical shifts does not required any randomization as soon as the protein is sufficiently large (number of residues > 50) to be well structured. For proteins or peptides of weaker size, the randomization factor increases with the number of residues. Exceptions occur:

(1) When the protein is not well-structured. (As in echistatine where there are nearly no interpeptidic hydrogen bond); for each proton the chemical average environment better remembers the random coil state and the randomization factor is small.
(2) When the structure is impeded (for instance by tight disulphide bridges in the AVP-like monomer) the chemical environment of a given proton is not averaged and remains specific of the 3D structure; the randomization factor is greater than expected.

Deriving Constraints From Chemical Shifts

The energy of the 50 refined conformers of the AVP-like, parallel dimer, which were defined from 62 NOE constraints is stretched over a range of 30 kcal (Figure 1). The percentage of NOE violations remains the same throughout the different conformers (about 15% of violations greater than 0.20 Å). Usually H_α

Table 4 The Randomization Factor for Small Peptides and Proteins[c]

Name	Pdb code	nb residues	p	Average chemical shift error[a] on		
				H_N	H_α	$H_{side\text{-}chain}$
Ribonuclease	3rn3(RX)	124	1.10	0.384	0.250	0.191
Tendamistat	3ait	74	1.00	0.397	0.281	0.186
Neurotoxin	1ntx	62	1.10	0.430	0.353	0.201
Trypsin inhibitor	1pit	58	1.00	0.391	0.249	0.184
ovomucoid trypsin inhibitor	1 tur	56	1.00	0.382	0.253	0.166
Echistatin	2ech	49	0.50	0.358	0.208	0.136
Anemonetoxin	1atx	46	0.80	0.461	0.245	0.176
Charybdotoxin	2crd	37	0.90	0.325	0.240	0.159
Conotoxin	1cco	27	0.70	0.492	0.211	0.158
Endothelin	1edp	21	0.60	0.427	0.203	0.111
AVP-like pp dimer		18	0.45	0.078	0.067	0.049
AVP-like monomer (b)		9	0.65	0.170	0.069	0.059

[a] The average error is the average absolute difference between the computed and observed ^1H chemical shifts.
[b] To be published.
[c] p is the factor which gives the best fit between observed and calculated chemical shifts.All protein 3D structure are extracted from the Brookhaven Protein Data Bank [16].

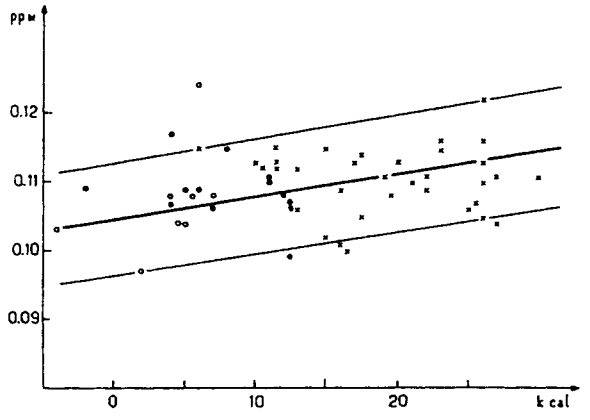

Figure 1 The distribution of the average error on the computed H_1 chemical shifts related to the relative energy of the conformers of the AVP-like parallel dimer. The top solutions are divided into two groups following the average error on the CysHβ protons. Open dots correspond to the weakest errors (0.100 p.p.m.) and filled dots to the largest ones (0.150 p.p.m.).

proton chemical shifts are largely dependent on the 3D structures of peptides and proteins (for instance, they are used to define the regular secondary structures in proteins [14]). The chemical shifts of the H_α protons were calculated for the 50 conformers using Eq. (2) with $p=0.45$. The error, expressed as the averaged absolute difference between each computed and observed H_α chemical shifts, ranged from 0.067 to 0.095 p.p.m. As illustrated in Figure 1, this average error steadily increases with the energy of the refined conformer and does not deviate by more than 0.01 p.p.m. from its mean value ($\langle dc \rangle = 0.105 + \Delta E \times 0.003$ (p.p.m.)). The average error on chemical shift defines the same best structures as the final constrainted energy.

In fact, the top solutions may be divided into two groups of 3D structures which deviate between them by more than 2.2 Å r.m.s. at the level of the 12 residue membered ring. Within each group the rms deviation is less than 0.8 [Å] from an average 3D structure. Both tripeptide tails are highly flexible and near a random coil conformation. The conformer energy, as well as the average error of the H_α chemical shift, does not allow one group of 3D structures to be favoured (see Figure 1).

On the contrary, a strong difference appears at the level of the disulphide bridges and is made obvious by the estimation of the chemical shifts of the eight cysteine H_β protons for each group of structures. For one group, the average error is about 0.100 p.p.m. whereas it is greater than 0.150 p.p.m. for the other. The use of NOE constraints for such a peptide allows a larger fluctuation of the solution structure than do the observed chemical shifts. The variation of the chemical shifts appears to be a better characteristic of the 3D structure of a flexible molecule than the NOE constraints. A refinement based on a back calculation of the chemical shifts would provide a better definition of the position of the different functional groups which involve these observed variations.

In NOE experiments, a correlation peak is observed as soon as two protons i and j are next to each other. The integrated volume of this peak is a function of the distance d_{ij} between both protons i and j (usually in $1/d_{ij}^6$). A constraint between i and j is then used during the classical energy refinement with a binomial holding function $k(d - d_{NOE})^2$. If the amount of NOE constraints is too small, large variations of expected solution structure may be spuriously involved.

On another way, chemical shifts provide a more accurate representation of the proton environment. They cannot be used in the first stage of the structure refinement because they harbour less specific information than the NOE constraints. The same chemical environment may be obtained from various manners, with not exactly the same surrounding atoms. Direct refinement based on the computation of proton chemical shift would suffer the same problem of numerous pseudo minima as an usual refinement based on van der Waals or solvatation energies involves. Constrainted refinement based on NOE distance constraints converges towards a limited number of solutions. When a set of approximate 3D structure has been generated from the available NOE constraints, the approximated variations of the chemical environment of each proton can be analysed. The environment among of the conformers which gives the best fit between the computed and observed chemical shift should correspond to the real chemical environment of that proton. Then from the different conformers, the best environment of each proton may be defined as a set of distances between this proton and different functional groups.

By this way, a second set of distance constraints may be derived and added to the initial NOE constraints in the energy refinement (with XPlor for instance). When all the available 3D structures were refined, the best chemical environments of each protons may be reactualized and the set of distance constraints optimized (and so on).

The search for the best environment of a given proton among a set of conformers allows the limita-

tion of errors which are involved through the existence of scarce (local) NOE constraints. As H_α proton chemical shifts mainly depend on the peptide 3D structure, the set of chemical shift-derived constraints is defined between these protons and neighbouring oxygen atoms for subsequent refining of the AVP-like parallel dimer. It quickly appears that an improvement of the peptide geometry controlled by the H_α chemical shift, also ensures an improvement of H_N and side chain proton chemical shifts.

The refinement was performed on the set of the five conformers with the best estimation of the Cys H_β protons. The initial mean errors in the estimation of proton chemical shifts were 0.082, 0.102 and 0.064 p.p.m. respectively, for H_N, H_α and side-chain protons; the final errors were respectively 0.074, 0.059 and 0.051 p.p.m. The initial average deviations between the conformers was 0.80 [Å] rms, but roughly remains the same between the final refined 3D structures.

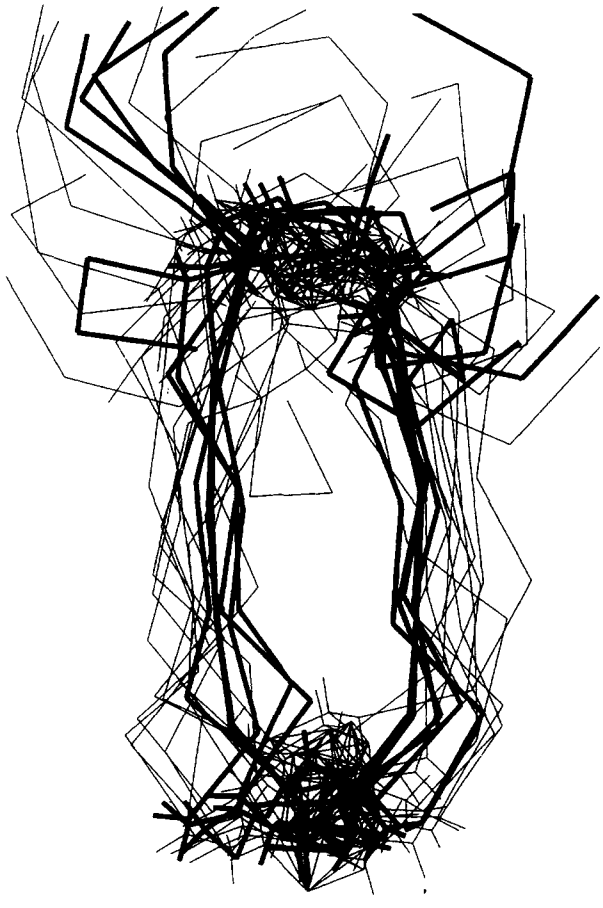

Figure 2 Superimposed solution structures of the AVP-like parallel dimer. The strong lines correspond to the conformers with the best conformation of the disulphide bridges, after refinement using both NOE and chemical-shift derived constraints.

Figure 2 illustrates the extra restriction of the solution structure of the AVP-like dimer, compared to the whole solutions which remain compatible with the NOE constraints.

CONCLUSIONS

For peptides and small proteins, the introduction of a randomization factor is required to ensure a good estimation of the chemical shifts from the 3D solution structure. Further investigations should be necessary to determine in what way this randomization factor informs about the mobility of the molecule.

We obtained for our dimeric peptide structure a better estimate of the observed chemical shifts than for other protein structures (Table 4). Hydrogen bonds and ring current effects which occur in all the proteins and are absent in the AVP-like dimer should be responsible for this observed discrepancy. They are not presently well-estimated and a more precise analysis of their influence is required. Undoubtedly the mobility of the different functional groups is important and should be taken into account in the estimation of the chemical shifts from 3D structures.

As already proved for proteins [7], the use of both chemical shift and NOE informations allows a best definition of the solution structure of a peptide. Short peptides should adopt a more defined structure in aqueous solutions [17] than is usually thought and their apparent flexibility can be strongly reduced by the use of all available informations.

ACKNOWLEDGEMENTS

We thank Dr J. Y. Lallemand and V. Stoven who allowed us to perform the NMR experiments at the Laboratoire de Synthèse Organique de l'Ecole Polytechnique (Palaiseau, France).

REFERENCES

1. K. Wüthrich (1989). Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science 243*, 45–50.
2. T. F. Havel (1986). DISGEO. Quantum Chemistry Program Exchange no. 507 Indiana University.
3. P. Güntert, W. Braun and K. Wüthrich (1991). Efficient computation of three dimensional protein structures in

solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J. Molec. Biol. 217*, 517–530.

4. Y. Kim and J. H. Prestegard (1990). Refinement of the NMR structures for acyl carrier protein with scalar coupling data. *Proteins: Struct., Funct. Genet. 8*, 377–385.
5. D. A. Mierke, A. Geyer and H. Kessler (1994). Coupling constants and hydrogen bonds as experimental restraints in a distance geometry refinement protocol. *Int. J. Peptide Protein Res. 44*, 325–331.
6. E. Oldfield (1995). Chemical shifts and three-dimensional protein structures. *J. Biomol. NMR 5*, 217–225.
7. E. Kuszewski, A. M. Gronenborg and G. M. Clore (1995). The impact of direct refinement against proton chemical shifts of protein structure determination by NMR. *J. Magn. Reson. 107*, 293–297.
8. H. Rik: *Nuclear Magnetic Resonance Spectroscopy. A Physical Chemical View*. Longman, Scientific and Technical, Harlow, UK, 1986.
9. H. Osapay and D. A. Case (1991). A new analysis of proton chemical shifts in protein. *J. Am. Chem. Soc. 13*, 9436–9444.
10. I. Herranz, C. Gonzalez, M. Rico, J. L. Nieto, J. Santoro, M. A. Jimenz, M. Bruix, J. L. Neira and F. J. Blanco (1992). Peptide group chemical shift computation. *Magn. Reson. Chem. 30*, 1012–1018.
11. A. Pardi, G. Wagner and K. Wüthrich (1983). Protein conformation and proton nuclear magnetic resonance chemical shifts. *Eur. J. Biochem. 137*, 445–454.
12. A. Bundi and K. Wüthrich (1979). 1H-NMR parameters of the common amino-acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-Ala-OH. *Biopolymers 18*, 285–297.
13. A. De Marco (1977). pH dependence of internal references. *J. Magn. Reson. 26*, 527–528.
14. D. S. Wishart, B. D. Sykes and F. M. Richards (1991). Relationship between nuclear magnetic resonance chemical shifts and protein secondary structure. *J. Molec. Biol. 222*, 311–333.
15. A. T. Brünger (1992). X-PLOR, Version 2.1 Manual. Yale University, New Haven, CT.
16. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, R. Rodgers, O, Kennard, T. Schimanouchi and M. Tasuml (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Molec. Biol. 112*, 535–542.
17. D. H. Dyson and P. E. Wright (1991). Defining solution conformations of small linear peptides. *Annu. Rev. Biophys. Biophys. Chem. 20*, 519–531.